

(19)



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11)

EP 1 672 528 A2

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
21.06.2006 Bulletin 2006/25

(51) Int Cl:
G06F 17/24 (2006.01)

(21) Application number: 05111549.1

(22) Date of filing: 01.12.2005

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI
SK TR
Designated Extension States:
AL BA HR MK YU

(72) Inventors:
• Rothschilder, Chad B.
98052, Redmond (US)
• Wu, Su-Piao B.
98052, Redmond (US)

(30) Priority: 20.12.2004 US 18912

(74) Representative: Grünecker, Kinkeldey,
Stockmair & Schwanhäusser
Anwaltssozietät
Maximilianstrasse 58
80538 München (DE)

(71) Applicant: MICROSOFT CORPORATION
Redmond, Washington 98052-6399 (US)

(54) File Formats, Methods, and Computer Program Products for Representing Workbooks

(57) File formats, methods, and computer program products are provided for representing a workbook in a modular content framework. The modular content framework may include a file format container associated with modular parts. A file format includes logically separate modular parts that are associated with each other by one or more relationships where each modular part is associated with a relationship type. The modular parts include a workbook part operative as a guide for properties of the workbook and a worksheet part associated with the

workbook part and operative to specify a definition of cells within a worksheet associated with the worksheet part. The modular parts may also include a document properties part containing built-in properties associated with the file format and a thumbnail part containing associated thumbnails. Each modular part is capable of being interrogated separately, extracted from the workbook, and/or reused in a different workbook.

EP 1 672 528 A2

Description

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This patent application is related to and filed with U.S. Patent Application, Attorney Docket No. 60001.0441US01, entitled "File Formats, Methods, and Computer Program Products For Representing Documents," filed on December 20, 2004; U.S. Patent Application, Attorney Docket No. 60001.0443US01, entitled "File Formats, Methods, and Computer Program Products For Representing Presentations," filed on December 20, 2004; and Attorney Docket No. 60001.0440US01, entitled "Management and Use of Data in a Computer-Generated Document," filed on December 20, 2004; all of which are assigned to the same assignee as this application. The aforementioned patent applications are expressly incorporated herein, in their entirety, by reference.

TECHNICAL FIELD

[0002] The present invention generally relates to file formats, and more particularly, is related to methods and file formats for representing features and data of workbooks in a componentized spreadsheet application program.

BACKGROUND

[0003] The Information age has facilitated an era of building informative spreadsheets utilizing spreadsheet software applications. However, the organization of features and data within previous spreadsheet file formats is very confusing and unclear to outside programmers and developers. For instance, previous spreadsheet file formats are created in the form of a single file using a binary record format containing all of the information required to render workbooks. Because proprietary formats are generally used to create these single files, writing code to work with and access these file formats without using the application program that created the file format is a nightmare for professional developers.

[0004] Another problem is basic document or worksheet re-use. For instance, it is very difficult to extract one or more worksheets from a workbook file and reuse the extracted worksheets in a different workbook and retain worksheet integrity, even in the same application. Comparatively, reusing worksheets between different applications is worse. Reusing content on a worksheet, for example reusing a table from EXCEL to WORD, is similarly difficult.

[0005] Additionally, because of the single file format, it is practically impossible to lock part of a workbook. Most of the technology in terms of file locking is all done at the file level, thus if a file is locked by a user, no other users can edit the file. Viewing is possible, but not editing.

[0006] There is also a problem of document interroga-

tion. Finding content within a workbook file, for example finding worksheets for a 2004 sales forecast, can be a daunting task. It is difficult to write code that programmatically finds cell A1 of a spreadsheet file and determines the contents of that cell (a string value, a formula, a calculated result) without using the same spreadsheet application that created the workbook. It is also very difficult to find parts of a single file format presentation and determine semantics about the content. For example, it is difficult to write code that programmatically locates a list of data in a spreadsheet application, and adds 3 rows of data to the list without using that spreadsheet application.

[0007] It is still difficult to implement reader and writer classes that can handle existing binary file formats well. Even if a tool targeted at an application was developed it could not interrogate all document formats. This problem is referred to as the opaqueness of single file formats.

[0008] Still further, due to intermingling of data, the ability to re-brand a worksheet, or multiple worksheets, is nearly impossible outside of the same spreadsheet application. Re-branding a worksheet involves taking a worksheet from workbook A, moving it to workbook B, and making the worksheet look as though it was authored in the normal authoring context of workbook B having the same text-based format.

[0009] Document surfacing, the ability to take pieces of a worksheet document and drop them into another document of a different application, is also a problem. For instance, a spreadsheet table copied into a presentation document is difficult to interrogate in the single file format.

[0010] Accordingly there is an unaddressed need in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

[0011] Embodiments of the present invention provide file formats, methods, and computer program products for representing a workbook in a modular content framework implemented within a computing apparatus. Embodiments of the present invention disclose an file format based on open standards, such as an extensible markup language (XML) file format and/or a binary file format, and a method by which features and data of a workbook are organized and modeled within a spreadsheet application file. The file format is designed such that it is made up of collections and parts. Each collection functions as a folder and each modular part functions as a file. These separate files are related together with relationships where each separate file is associated with a relationship type. This design greatly simplifies the way the spreadsheet applications organizes workbook features and data, and presents a logical model that is much less confusing.

One embodiment is a file format for representing a workbook in a modular content framework. The modular con-

tent framework may include a file format container associated with the modular parts. The file format includes modular parts that are logically separate but associated with one another by one or more relationships. Each modular part is associated with a relationship type and the modular parts include a workbook part operative as a guide for properties of the workbook. The modular parts also include a worksheet part associated with the workbook part and operative to specify a definition of cells within a worksheet, a sheet part containing data associated with a macro sheet, a chart sheet part containing data associated with defining a chart, and/or a dialog sheet part containing data associated with workbook dialog. Each modular part is capable of being interrogated separately with or without the spreadsheet application and without other modular parts being interrogated, which offers gains in efficiency when the workbook is queried.

[0012] The modular parts may also include a document properties part containing built-in properties associated with the file format and a thumbnail part containing one or more thumbnails associated with the file format. Each modular part is capable of being extracted from and/or copied from the workbook and reused in a different workbook along with associated modular parts identified by traversing the relationships of the modular part reused.

[0013] Another embodiment is a method for representing a workbook in a file format wherein modular parts associated with the workbook include each part written into the file format. The method involves writing a workbook part of the file format, querying the workbook for a worksheet relationship type, and writing a worksheet part of the file format separate from the workbook part. The method also involves establishing a relationship between the worksheet part and the workbook part. Additionally, the method may involve establishing a relationship between the workbook part and a file format container where the file format container includes a document properties part containing built-in properties associated with the file format and a thumbnail part containing a thumbnail associated with the file format.

[0014] Still further, the method may involve writing modular parts associated with relationship types wherein the modular parts that are to be shared are written only once, and establishing relationships to the modular parts written. Writing the modular parts may also involve examining data associated with the workbook, determining whether the data examined has been written to a modular part, and when the data examined has not been written to the modular part, writing the modular part to include the data examined.

[0015] Still another embodiment is a computer program product including a computer-readable medium having control logic stored therein for causing a computer to represent a workbook in a file format where modular parts of the file format include each part written into the file format. The control logic includes computer-readable program code for causing the computer to write a work-

book part of the file format, query the workbook for a worksheet relationship type, write a worksheet part of the file format separate from the workbook part, and establish a relationship between the worksheet part and the workbook part.

[0016] The invention may be implemented utilizing a computer process, a computing system, or as an article of manufacture such as a computer program product or computer readable media. The computer program product may be a computer storage media readable by a computer system and encoding a computer program of instructions for executing a computer process. The computer program product may also be a propagated signal on a carrier readable by a computing system and encoding a computer program of instructions for executing a computer process.

[0017] These and various other features, as well as advantages, which characterize the present invention, will be apparent from a reading of the following detailed description and a review of the associated drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018]

FIGURE 1 is a computing system architecture illustrating a computing apparatus utilized in and provided by various illustrative embodiments of the invention;

FIGURES 2a-2c are block diagrams illustrating a workbook relationship hierarchy for various modular parts utilized in a file format for representing a workbook according to various illustrative embodiments of the invention; and

FIGURES 3-4 are illustrative routines performed in representing workbooks in a modular content framework according to illustrative embodiments of the invention.

DETAILED DESCRIPTION

[0019] Referring now to the drawings, in which like numerals represent like elements, various aspects of the present invention will be described. In particular, FIGURE 1 and the corresponding discussion are intended to provide a brief, general description of a suitable computing environment in which embodiments of the invention may be implemented. While the invention will be described in the general context of program modules that execute in conjunction with program modules that run on an operating system on a personal computer, those skilled in the art will recognize that the invention may also be implemented in combination with other types of computer systems and program modules.

[0020] Generally, program modules include routines, programs, operations, components, data structures, and other types of structures that perform particular tasks or implement particular abstract data types. Moreover,

those skilled in the art will appreciate that the invention may be practiced with other computer system configurations, including hand-held devices, multiprocessor systems, microprocessor-based or programmable consumer electronics, minicomputers, mainframe computers, and the like. The invention may also be practiced in distributed computing environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules may be located in both local and remote memory storage devices.

[0021] Referring now to FIGURE 1, an illustrative computer architecture for a computer 2 utilized in an embodiment of the invention will be described. The computer architecture shown in FIGURE 1 illustrates a computing apparatus, such as a server, desktop, laptop, or handheld computing apparatus, including a central processing unit 5 ("CPU"), a system memory 7, including a random access memory 9 ("RAM") and a read-only memory ("ROM") 11, and a system bus 12 that couples the memory to the CPU 5. A basic input/output system containing the basic routines that help to transfer information between elements within the computer, such as during startup, is stored in the ROM 11. The computer 2 further includes a mass storage device 14 for storing an operating system 16, application programs, and other program modules, which will be described in greater detail below.

[0022] The mass storage device 14 is connected to the CPU 5 through a mass storage controller (not shown) connected to the bus 12. The mass storage device 14 and its associated computer-readable media provide non-volatile storage for the computer 2. Although the description of computer-readable media contained herein refers to a mass storage device, such as a hard disk or CD-ROM drive, it should be appreciated by those skilled in the art that computer-readable media can be any available media that can be accessed by the computer 2.

[0023] By way of example, and not limitation, computer-readable media may comprise computer storage media and communication media. Computer storage media includes volatile and non-volatile, removable and non-removable media implemented in any method or technology for storage of information such as computer-readable instructions, data structures, program modules or other data. Computer storage media includes, but is not limited to, RAM, ROM, EPROM, EEPROM, flash memory or other solid state memory technology, CD-ROM, digital versatile disks ("DVD"), or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by the computer 2.

[0024] According to various embodiments of the invention, the computer 2 may operate in a networked environment using logical connections to remote computers through a network 18, such as the Internet. The computer 2 may connect to the network 18 through a network in-

terface unit 20 connected to the bus 12. It should be appreciated that the network interface unit 20 may also be utilized to connect to other types of networks and remote computer systems. The computer 2 may also include an input/output controller 22 for receiving and processing input from a number of other devices, including a keyboard, mouse, or electronic stylus (not shown in FIGURE 1). Similarly, an input/output controller 22 may provide output to a display screen, a printer, or other type of output device.

[0025] As mentioned briefly above, a number of program modules and data files may be stored in the mass storage device 14 and RAM 9 of the computer 2, including an operating system 16 suitable for controlling the operation of a networked personal computer, such as the WINDOWS XP operating system from MICROSOFT CORPORATION of Redmond, Washington. The mass storage device 14 and RAM 9 may also store one or more program modules. In particular, the mass storage device 14 and the RAM 9 may store a spreadsheet application program 10. The spreadsheet application program 10 is operative to provide functionality for the creation and structure of workbooks, such as a workbook 27, in an open file format 24, such as an XML file format. According to one embodiment of the invention, the spreadsheet application program 10 and other application programs 26 comprise the OFFICE suite of application programs from MICROSOFT CORPORATION including the WORD, EXCEL, and POWERPOINT application programs.

[0026] Embodiments of the present invention greatly simplify and clarify the organization of workbook features and data. The spreadsheet program 10 organizes the 'parts' of a workbook file (features, data, themes, styles, objects, etc) into logical, separate pieces, and then expresses relationships among the separate parts. These relationships, and the logical separation of 'parts' of a workbook, make up a new file organization that can be easily accessed, such as by a developer's code, without using the spreadsheet application itself.

[0027] Referring now to FIGURES 2a-2c, block diagrams illustrating a workbook relationship hierarchy for various modular parts utilized in the file format 24 for representing a workbook according to various illustrative embodiments of the invention will be described. The workbook relationship hierarchy 208 lists specific spreadsheet application relationships. Optional relationships with respect to validation are indicated in italics, and dashed connecting lines 203 indicate a one to potentially many relationship. Thus, for example there is a worksheet part 217 for each worksheet associated with a workbook 202.

[0028] The various modular parts or components of the presentation hierarchy 208 are logically separate but are associated by one or more relationships. Each modular part is also associated with a relationship type and is capable of being interrogated separately with or without the spreadsheet application program 10 and/or with or without other modular parts being interrogated. Thus, for

example, it is easier to locate the contents of a worksheet cell because instead of searching through all the binary records for cell information, code can be written to easily inspect the relationships in a workbook and find the worksheet parts, effectively ignoring the other features and data in the file format 24. Thus, the code is written to step through the cells in a much simpler fashion than previous interrogation code. Further, 'authoring' scenarios, where a developer writes code to insert a new part, or to insert a completely new file without running the spreadsheet application, are simplified due to the modular part file format.

[0029] A modular content framework may include a file format container 204 associated with the modular parts. The modular parts include, the workbook part 202 operative as a guide for properties of the workbook and the worksheet part 217 associated with the workbook part 202 and operative to specify a definition of cells within a worksheet associated with the worksheet part 217. The workbook hierarchy 208 may also include a document properties part 205 containing built-in properties associated with the file format 24, and a thumbnail part 207 containing a thumbnail associated with the file format 24. [0030] The modular parts also include a sheet part 210 containing data associated with a macro sheet, a chart sheet part 212 containing data associated with defining a chart, and a style sheet part 240 representing a theme in the workbook. It should be appreciated that each modular part is capable of being extracted from or copied from the workbook and reused in a different workbook along with associated modular parts identified by traversing relationships of the modular part reused. Associated modular parts are identified when the spreadsheet application 10 traverses inbound and outbound relationships of the modular part reused.

[0031] Other modular parts may include a style part 220 containing data associated with a style at a cell level in the workbook, a dialog sheet part 214 containing data associated with workbook dialog, a markup maps part 218 containing visuals depicting a markup language format associated with the workbook, and a shared strings part 222 containing a string associated with a plurality of cells in the workbook. Still other modular parts include a workbook connections part 224 containing data associated with interfacing with the workbook, a background picture part 225, a mail envelope part 242 containing envelope data where a user of the workbook has sent the workbook via electronic mail, a code file part 244 containing code associated with the workbook, and a comments part 247 containing comments associated with the workbook.

[0032] Still further, the modular parts may include a schemas part 254 containing schemas associated with the markup maps part 218, an image part 248 containing image data associated with the workbook, and an embedded object part 230 containing an object associated with the workbook. Other modular parts may also include a user data part 245 containing customized data capable

of being read into the workbook and changed, a drawing object part 257 containing an object built using a drawing platform, a legacy drawing object part 252, such as an Escher 1.0 object, a table index part 232 containing data defining a table index associated with the worksheet, and a list part 228 containing data defining a list associated with the worksheet. As an example, embodiments of the present invention make it easier to locate a list in a workbook because any list has a list part 228 separate in the file format 24 with corresponding relationships expressed. The list part 228, as are other modular parts, is logically broken-out and separate from other features & data of the workbook. Further, because the logical structure of a list is clearly understood, it is also less complicated to add more rows of data to a list.

[0033] Other modular parts associated with the workbook may include a pivot table part 234 containing data defining a pivot table associated with the worksheet, a pivot cache definition part 235 containing data defining a cache associated with the pivot table, and a pivot cache records part 237 containing data associated with the pivot cache definition part. A pivot table is a program tool that allows selected columns and rows of data in a spreadsheet or database table to be reorganized and summarized in order to obtain a desired report. A pivot table turns the data to view it from different perspectives. It should be appreciated that modular parts that are shared in more than one relationship are typically only written to the file once. It should also be appreciated that certain modular parts are global and thus, can be used anywhere in the file format. In contrast, some modular parts are non-global and thus, can only be shared on a limited basis.

[0034] In various embodiments of the invention, the file format 24 may be formatted according to extensible markup language ("XML") and/or a binary format. As is understood by those skilled in the art, XML is a standard format for communicating data. In the XML data format, a schema is used to provide XML data with a set of grammatical and data type rules governing the types and structure of data that may be communicated. The XML data format is well-known to those skilled in the art, and therefore not discussed in further detail herein. The XML formatting closely reflects the internal memory structure of an entire workbook. Thus, an increase in load and save speed is evident.

[0035] Embodiments of the present invention make workbooks more programmatically accessible. This enables a significant number of new uses that are simply too hard for previous file formats to accomplish. For instance, utilizing embodiments of the present invention, a server-side program is able to create a workbook for someone based on their input. For example, creating an analysis report on Company A for the time period of 1/1/2004-12/31/2004 where all variable input is italicized.

[0036] Other examples include, an external process scanning and rewriting all workbooks on a network in order to update a company logo and visual color scheme,

a chart in one workbook being programmatically copied and inserted into another workbook, and calculation results being retrieved from a workbook and updated to a database nightly. Still another example is a government agency can process workbooks, and more easily convert the features and data to their internal text-based format.

[0037] FIGURES 2a-2c also include relationship types utilized in the file format 24 according to various illustrative embodiments of the invention. The relationship types associated with the modular parts not only identify an association or dependency but also identify the basis of the dependency. The relationship types include the following: a code file relationship capable of identifying code files, a user data relationship, a style sheet relationship, a comments relationship, an embedded object relationship, a drawing object relationship, an image relationship, a sound relationship, a mail envelope relationship, a document properties relationship, a thumbnail relationship, a schema relationship, a chart sheet relationship, a dialog sheet relationship, a worksheet relationship, and a pivot table relationship, a shared string relationship, a lists relationship, a pivot cache relationship, a styles relationship, a markup maps relationship, and a pivot cache metadata relationship.

[0038] Referring to FIGURE 2c also illustrates the listing 259 that lists collection types for organizing the modular parts. The collection types include a chart sheet collection including the chart sheet part 212, a dialog sheet collection including the dialog sheet part 214, a worksheets collection including the worksheet part 217, a pivots collection including the pivot table part 234, and a pivot cache collection including the pivot cache definition part 235 and the pivot cache records part 237. The collection types also include a styles collection including the styles sheet part 240 and the styles part 220, a markup maps collection including the markup maps part 218, a lists collection including the lists part 228, and an embeddings collection including the embedded object part 230 and the user data part 245.

[0039] FIGURES 3-4 are illustrative routines performed in representing workbooks in a modular content framework according to illustrative embodiments of the invention. When reading the discussion of the routines presented herein, it should be appreciated that the logical operations of various embodiments of the present invention are implemented (1) as a sequence of computer implemented acts or program modules running on a computing system and/or (2) as interconnected machine logic circuits or circuit modules within the computing system. The implementation is a matter of choice dependent on the performance requirements of the computing system implementing the invention. Accordingly, the logical operations illustrated in FIGURES 3-4, and making up the embodiments of the present invention described herein are referred to variously as operations, structural devices, acts or modules. It will be recognized by one skilled in the art that these operations, structural devices, acts and modules may be implemented in software, in

firmware, in special purpose digital logic, and any combination thereof without deviating from the spirit and scope of the present invention as recited within the claims set forth herein.

[0040] Referring now to FIGURES 2a-2c and 3, the routine 300 begins at operation 304, where the spreadsheet application program 10 writes the workbook part 202. The routine 300 continues from operation 304 to operation 305, where the spreadsheet application program 10 queries the workbook for worksheet relationships. Next, at operation 307, the spreadsheet application writes the worksheet parts 217 referenced in the workbook part 202 and establishes relationships between each worksheet part 217 and the workbook part 202.

[0041] Next, at operation 308, the spreadsheet application 10 writes other modular parts associated with relationship types, such as the image part 248, and the schema part 254. Any modular part to be shared between other modular parts is written only once. The routine 300 then continues to operation 310.

[0042] At operation 310, the spreadsheet application 10 establishes relationships between newly written and previously written modular parts. The routine 300 then terminates at return operation 312.

[0043] Referring now to FIGURE 4, the routine 400 for writing modular parts will be described. The routine 400 begins at operation 402 where the spreadsheet application 10 examines data in the spreadsheet application. The routine 400 then continues to detect operation 404 where a determination is made as to whether the data has been written to a modular part. When the data has not been written to a modular part, the routine 400 continues from detect operation 404 to operation 405 where the spreadsheet application writes a modular part including the data examined. The routine 400 then continues to detect operation 407 described below.

[0044] When at detect operation 404, if the data examined has been written to a modular part, the routine 400 continues from detect operation 404 to detect operation 407. At detect operation 407 a determination is made as to whether all the data has been examined. If all the data has been examined, the routine 400 returns control to other operations at return operation 412. When there is still more data to examine, the routine 400 continues from detect operation 407 to operation 410 where the spreadsheet application 10 points to other data. The routine 400 then returns to operation 402 described above.

[0045] Based on the foregoing, it should be appreciated that the various embodiments of the invention include file formats, methods and computer program products for representing workbooks in a modular content framework. The above specification, examples and data provide a complete description of the manufacture and use of the composition of the invention. Since many embodiments of the invention can be made without departing from the spirit and scope of the invention, the invention

resides in the claims hereinafter appended.

Claims

1. A file format for representing a workbook within a spreadsheet application, the file format representing the workbook in a modular content framework implemented within a computing apparatus, the file format comprising:

modular parts logically separate but associated by one or more relationships wherein each modular part is associated with a relationship type and wherein the modular parts include:

a workbook part operative as a guide for properties of the workbook; and at least one of the following:

a worksheet part associated with the workbook part and operative to specify a definition of cells within a worksheet associated with the worksheet part;
a sheet part containing data associated with a macro sheet;
a chart sheet part containing data associated with defining a chart; and
a dialog sheet part containing data associated with workbook dialog;

wherein each modular part is capable of being interrogated separately without the spreadsheet application and without other modular parts being interrogated.

2. The file format of claim 1, wherein the modular content framework includes a file format container associated with the modular parts wherein the modular parts further include:

a document properties part containing built-in properties associated with the file format; and
a thumbnail part containing thumbnails associated with the file format.

3. The file format of claim 1, wherein each modular part is capable of being one of extracted from and copied from the workbook and reused in a different workbook along with associated modular parts identified by traversing relationships of the modular part reused.

4. The file format of claim 3, wherein the modular parts further include at least one of the following:

a style sheet part representing a theme in the workbook;
a style part containing data associated with a

style at a cell level in the workbook;
a comments part containing comments associated with the workbook;
a markup maps part containing visuals depicting a markup language format associated with the workbook;
a schemas part containing schemas associated with the markup maps part;
a shared strings part containing a string associated with a plurality of cells in the workbook;
a workbook connections part containing data associated with interfacing with the workbook;
a mail envelope part containing envelope data where a user of the workbook has sent the workbook via electronic mail; and
a code file part containing code associated with the workbook.

5. The file format of claim 4, wherein the modular parts further include at least one of the following:

an image part containing image data associated with the workbook;
an embedded object part containing an object associated with the workbook;
a user data part containing customized data capable of being read into the workbook and changed; and
a drawing object part containing an object built using a drawing platform.

6. The file format of claim 5, wherein the modular parts further include at least one of the following:

a table index part containing data defining a table index associated with the worksheet;
a list part containing data defining a list associated with the worksheet;
a pivot table part containing data defining a pivot table associated with the worksheet;
a pivot cache definition part containing data defining a cache associated with the pivot table; and
a pivot cache records part containing data associated with the pivot cache definition part.

7. The file format of claim 6, wherein at least some of the modular parts are organized in collection types and wherein the collection types include at least one of the following:

a chart sheet collection including the chart sheet part;
a dialog sheet collection including the dialog sheet part;
a worksheets collection including the worksheet part;
a pivots collection including the pivot table part;

- a pivot cache collection including at least one of the pivot cache definition part and the pivot cache records part;
 a styles collection including at least one of the styles sheet part and the styles part;
 a markup maps collection including the markup maps part;
 a lists collection including the lists part; and
 an embeddings collection including at least of the embedded object part and the user data part.
8. The file format of claim 3, where the relationship types associated with the modular parts comprises at least one of a code file relationship capable of identifying code files, a user data relationship, a hyperlink relationship, a styles sheet relationship, a comments relationship, an embedded object relationship, a drawing object relationship, an image relationship, a mail envelope relationship, a document properties relationship, a thumbnail relationship, a schema relationship, a chart sheet relationship, a dialog sheet relationship, a worksheet relationship, and a pivot table relationship, a shared string relationship, a lists relationship, a pivot cache records relationship, a styles relationship, a markup maps relationship, and a pivot cache metadata relationship.
9. The file format of claim 3, wherein content of the worksheet is capable of being one of extracted from and copied from the workbook and reused in a different workbook.
10. The file format of claim 3, wherein each modular part is capable of being locked separately while the other modular parts remain available for locking whereby multiple editors may each concurrently edit a modular part of the file format.
11. The file format of claim 3, wherein the modular parts are capable of providing semantics about content within the workbook when a modular part is interrogated.
12. The file format of claim 3, wherein at least one of the modular parts is authored in an authoring context of the workbook and wherein the at least one modular part in the workbook is capable of being one of extracted from and copied from the workbook and moved to a different workbook and wherein the at least one modular part is further capable of being altered to appear as though the at least one modular part was authored in an authoring context of the different workbook.
13. The file format of claim 3, wherein the file format is capable of providing a high-resolution thumbnail preview of each worksheet in the workbook.
14. The file format of claim 1, wherein the file format is formatted according to at least one of a markup language format and a binary format.
15. A method for representing a workbook in a file format wherein modular parts associated with the workbook include each part written into the file format, the method comprising:
 writing a workbook part of the file format;
 querying the workbook part for a worksheet relationship type;
 writing a worksheet part of the file format separate from the workbook part; and
 establishing a relationship between the worksheet part and the workbook part.
16. The method of claim 15, further comprising establishing a relationship between the workbook part and a file format container wherein the file format container includes at least one of the following:
 a document properties part containing built in properties associated with the file format; and
 a thumbnail part containing a thumbnail associated with the file format.
17. The method of claim 15, further comprising:
 writing other modular parts associated with relationship types wherein the other modular parts that are to be shared are written only once; and
 establishing relationships to the other modular parts written.
18. The method of claim 17, wherein writing the other modular parts associated with the relationship types comprises:
 a) examining data associated with the workbook;
 b) determining whether the data examined has been written to a modular part;
 c) when the data examined has not been written to the modular part, writing the modular part to include the data examined, examining other data associated with the workbook, and repeating b) through d); and
 d) when the data examined has been written to the modular part, examining other data associated with the workbook and repeating b) through d).
19. A computer program product comprising a computer-readable medium having control logic stored therein for causing a computer to represent a workbook in a file format comprising modular parts wherein the modular parts of the file format include each

part written into the file format, the control logic comprising computer-readable program code for causing the computer to:

write a workbook part of the file format; 5
query the workbook part for a worksheet relationship type;
write a worksheet part of the file format separate from the workbook part; and
establish a relationship between the worksheet 10
part and the workbook part.

20. The computer program product of claim 19, wherein the file format comprises at least one of a markup 15
file format written in a markup language and a binary format.

20

25

30

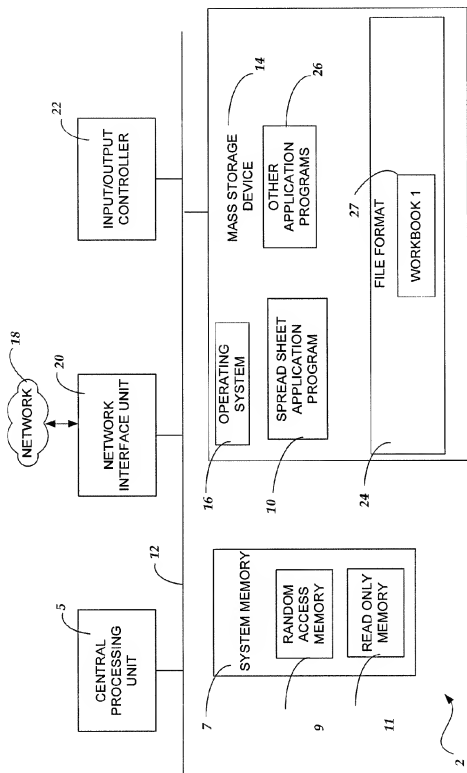
35

40

45

50

55



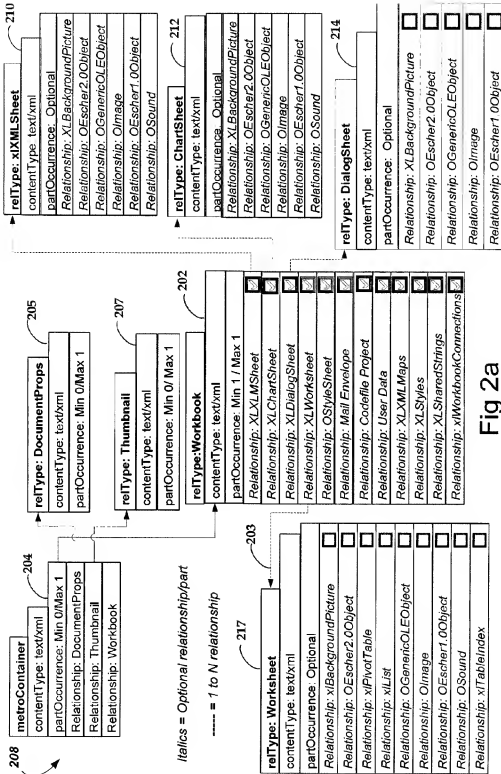


Fig 2a

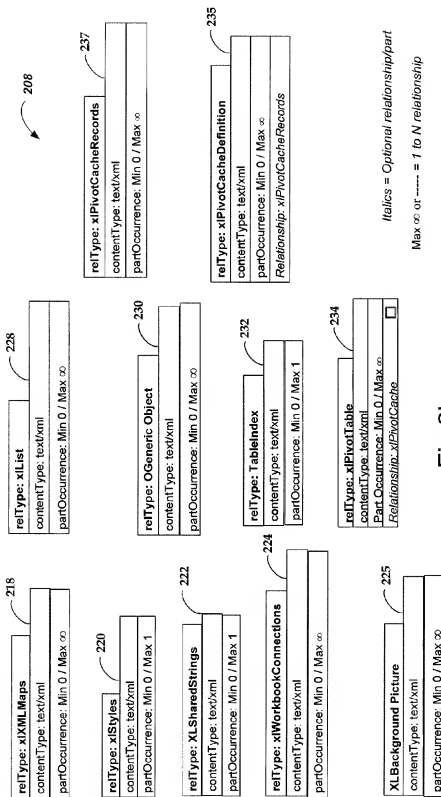


Fig 2b

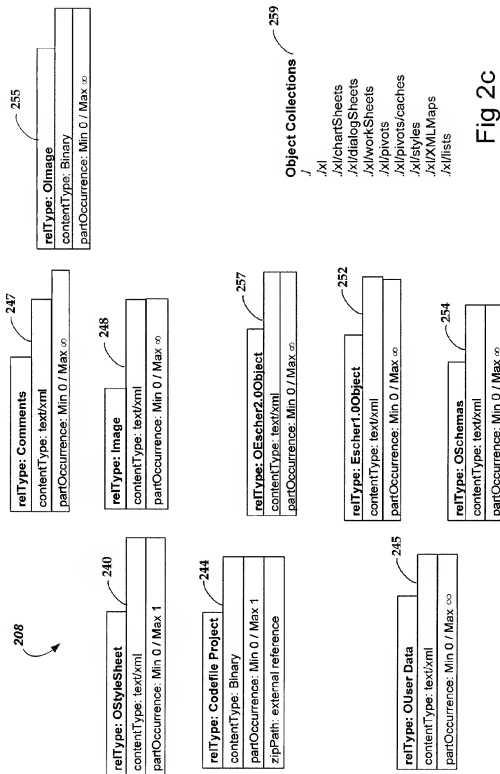


Fig 2c

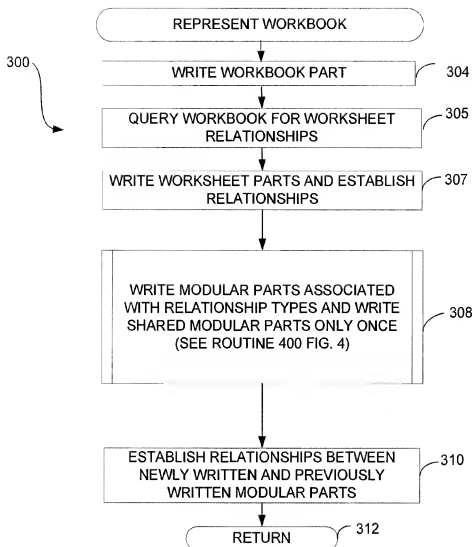


Fig. 3

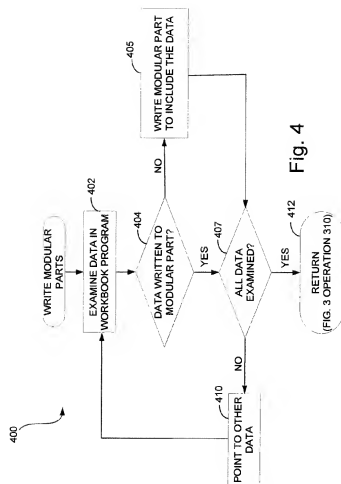


Fig. 4